EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Water surface object detection using panoramic vision based on improved single-shot multibox detector

Aofeng Li[1], Xufang Zhu[1*], Shuo He[1] and Jiawei Xia[2]

*Correspondence: 1580284687@qq.com
[1] College of Electronic Engineering, Naval University of Engineering, Wuhan 430000, China
Full list of author information is available at the end of the article

Abstract

In view of the deficiencies in traditional visual water surface object detection, such as the existence of non-detection zones, failure to acquire global information, and deficiencies in a single-shot multibox detector (SSD) object detection algorithm such as remote detection and low detection precision of small objects, this study proposes a water surface object detection algorithm from panoramic vision based on an improved SSD. We reconstruct the backbone network for the SSD algorithm, replace VVG16 with a ResNet-50 network, and add five layers of feature extraction. More abundant semantic information of the shallow feature graph is obtained through a feature pyramid network structure with deconvolution. An experiment is conducted by building a water surface object dataset. Results showed the mean Average Precision (mAP) of the improved algorithm are increased by 4.03%, compared with the existing SSD detecting Algorithm. Improved algorithm can effectively improve the overall detection precision of water surface objects and enhance the detection effect of remote objects.

Keywords: Panoramic vision, SSD, Deconvolution, Feature pyramid

1 Introduction

In recent years, the status of water transportation has been continuously improved in the field of transportation. As the main tool of water transportation, ships have received extensive attention from all walks of life for their safe, green and efficient operations. The development of emerging technologies such as artificial intelligence, the Internet, and big data has set off a research boom in smart ships [1]. As an important component of intelligent ship environment perception, surface object detection technology is the prerequisite and foundation for unmanned and intelligent ships, and has gradually become a new hot spot in the current intelligent ship research field. In ocean navigation, the automatic detection of surface objects is of great significance for the distribution of surface vessels, effective management of ship parking, identification of information on passing vessels, and realization of automatic collision avoidance.

With the continuous development of deep learning technology, object detection algorithms based on convolutional neural networks (CNN) have been proposed one after another, CNN is a kind of deep feedforward neural network, which has many

successful applications in image classification, object detection, object segmentation and many image and video domains.Water surface object detection technology is starting to adopt this method with high accuracy, fast speed and strong generalization ability. Surface object detection requires high real-time performance and recognition accuracy, and the monitored object needs to be detected at a long distance. Although the actual length of the long distance water surface object can reach tens of meters or even hundreds of meters, it only occupies dozens or even fewer pixels on the imaging plane. Among the current various object detection algorithms, the Single-shot Multibox Detector (SSD) algorithm has a good performance in detection speed and detection accuracy. However, its detection accuracy for small objects is not as satisfactory as expected, when the SSD algorithm is applied to the detection of water surface objects, it is difficult to detect small objects in the long distance water surface. Water surface object detection technology is starting to adopt this method with high accuracy, fast speed and strong generalization ability. Surface object detection requires high real-time performance and recognition accuracy, and the monitored object needs to be detected at a long distance. Although the actual length of the long distance water surface object can reach tens of meters or even hundreds of meters, it only occupies dozens or even fewer pixels on the imaging plane. Among the current various object detection algorithms, the Single-shot Multibox Detector(SSD) algorithm has a good performance in detection speed and detection accuracy. However, its detection accuracy for small objects is not as satisfactory as expected, when the SSD algorithm is applied to the detection of water surface objects, it is difficult to detect small objects in the long distance water surface.

Currently, object detection is mostly based on traditional vision. Traditional vision detection system have many deficiencies, such as limited vision, existence of non-detection zones, and the failure to acquire global information. A panoramic visual image is a composite image with high resolution and a wide viewing angle obtained by processing a few overlapped images. It has advantages such as a fast generation rate, high resolution, high fidelity of scene restoration, and low hardware requirements [2]. The unique advantages of the panoramic vision system can solve the viewing angle limitations of the traditional vision system, and more effectively meet the needs of large field of view, large range, and long distance. It is widely used in the field of intelligent navigation of ships [3].

In this work, a panoramic camera that generates a panoramic image will be used as a detection tool, use an improved SSD algorithm to detect water surface objects.

The main contributions of the work are as follows:

(1) Preprocess the panoramic image to obtain a rectangular panoramic image.
(2) Design a new SSD algorithm backbone network to obtain rich small object semantic information.
(3) Adopt Feature Pyramid Network (FPN) structure and add deconvolution operation to solve the problem of the lack of semantic information in the shallow feature map and the lack of detailed information in the deep feature map of the SSD model.
(4) Construct a water surface object dataset to verify the effectiveness of the method.

The remainder of this paper is organized as follows. Section 2 describes previous work in object detection. Section 3 introduced the improvement method we proposed. Section 4 describes dataset, training environment and evaluation indicators of the system. Section 5 reports and discusses the experimental results. Section 6 presents the conclusions of this paper.

2 Related work

Surface object detection methods are mainly divided into traditional detection methods and deep learning detection methods. Traditional object detection methods are based on background modeling, image segmentation, feature extraction and learning, color features, and color space transformation, or saliency detection [4]. For instance, Wijnhoven et al. [5] used the histogram of oriented gradients (HOG) to extract ship object features, and achieve object recognition through online learning and design classifiers. Mirghasemi et al. [6]. used the particle swarm optimization algorithm to transform the color space of the object to improve the robustness of the object recognition algorithm. Albrecht et al. [7] achieved saliency detection by constructing regional complexity features, regional differences features, surrounding differences features, and water and sky classification methods to improve the detection effect of the saliency analysis algorithm. These methods conduct object classification and detection for specific tasks; hence they are characterized by poor generalization ability and long detection times.

Object detection methods based on deep learning mainly include one-stage and two-stage methods. The one-stage object detection method uses the idea of regression analysis, omits the stage of candidate region generation, and directly obtains object classification and location information. The one-stage methods include the You Only Look Once (YOLO) [8–11] series and single-shot multibox detector (SSD) [12] series, which have fast but less accurate. The two-stage object detection method mainly generates candidate regions through selective search or region proposal network (RPN), and then perform classification and regression on the candidate regions to obtain the detection results. The two-stage methods include Fast-RCNN, Faster-RCNN, R-FCN, and Mask-RCNN [13–16], which have high accuracy but slow speed. To solve the ship identification problem in a video image, Cao et al. [17] proposed to extract object features based on image segmentation and convolutional neural networks and realized automatic identification of ships. Qi et al. [18] improved Faster R-CNN, and implemented ship object detection with image downscaling and scene narrowing methods, which shortened the detection time and improved the detection precision of Faster R-CNN. Lin et al. [19] proposed a new network architecture based on the faster R-CNN is proposed to further improve the detection performance by using squeeze and excitation mechanism. However, using squeeze and excitation mechanism will cause a decrease in the real-time performance of the network. According to the characteristics of ship target in the SAR images, Xiao et al. [20] make several improvements such as enlarging the input, proposal optimization, database target categorization, and weight balance on the basis of the standard Faster R-CNN. Zhang et al. [21] used two deep learning algorithms, the Mask R-CNN algorithm and the Faster R-CNN algorithm to build ship target feature extraction and recognition models based on deep convolutional neural networks.The above is based on the two-stage detection method, the detection accuracy can be high

Table 1 Comparison of various water surface object detection algorithms

Object detection algorithm	Algorithm characteristics	Advantage	Limitation
Improve Faster-RCNN[18–20]	Image downscaling and scene narrowing [18] Squeeze and excitation echanism[19] Enlarging the inputproposal optimization, database target categorization, and weight balance [20]	High accuracy	Low real-time
Mask-RCNN [20]	Instance segmentation [21]	High accuracy	Low real-time
Improve YOLO [22–25]	Reconstructed the shallow information and introduced a residual network [22] Designed an adaptive feature fusion module and new loss function [23] Fused DenseNet in YOLOV3 [24] Improved YOLOv3 and Real-time tracking [25]	High real-time	Low accuracy
Improve SSD [28]	Lightweight feature optimizing network and feature fusion module [28]	High accuracy and real-time	No object classification
our	FPN structure, deconvolution operation and panoramic vision	Wider range, High accuracy	Real-time declined slightly

and satisfactory, but the real-time performance is low. Li et al. [22] proposed a ship object detection algorithm based on the improved YOLOV3-Tiny, intensified and reconstructed the shallow information based on characteristics of ship objects, and introduced a residual network that greatly improved the accuracy of ship object detection on the sea surface. Xu et al. [23] proposed a YOLO-based multiscale object detection algorithm and designed an adaptive feature fusion module and new loss function. This algorithm has obvious advantages in terms of multiscale detection, for it improves the object detection precision without increasing the detection time. Li et al. [24] proposed a novel target detection method by fusing DenseNet in YOLOV3 to improve the stability of detection to decrease the feature loss, while the target feature is transmitted in the layers of a deep neural network. Jie et al. [25] presented ship detection and tracking of ships using the improved YOLOv3 detection algorithm and Deep Simple Online and Real-time Tracking (Deep SORT) tracking algorithm.

Compared with Faster R-CNN and YOLO, the SSD algorithm has integrated grid thinking from the YOLO algorithm and an anchor mechanism from Faster R-CNN. As a result, the SSD algorithm can quickly detect objects without decreasing the detection accuracy. Li et al. [26] applied the SSD algorithm to a railway scene with UAV surveillance, designed a three-step, multi-block SSD mechanism, and conducted sample detection with the shift learning method, which overall accuracy increased by 9.2% over traditional SSD. Yin et al. [27] formulated an object detection algorithm classifying and extracting the multiscale feature graph, dividing feature graphs with different scales in the SSD algorithm into low- and high-level feature graphs. The low-level feature graph extracts features of a shallow feature enhancement (SFE) module, and the high-level feature graph adopts two-stage deconvolution, which increases mAP by 2.4% compared to the SSD algorithm. Zhang et al. [28] proposed a lightweight feature optimizing network

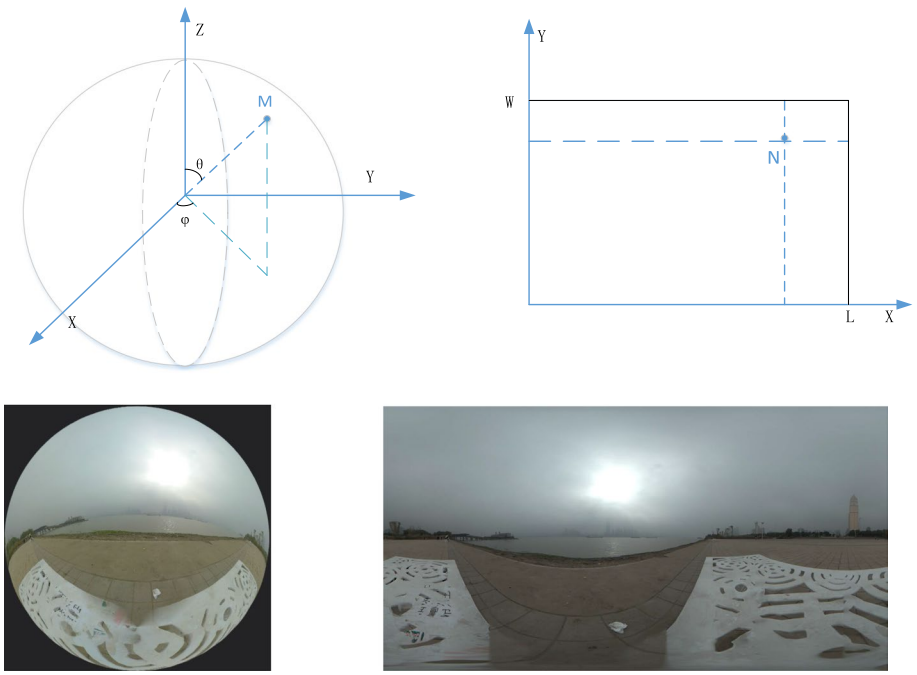

Fig.1 Coordinate conversion diagram

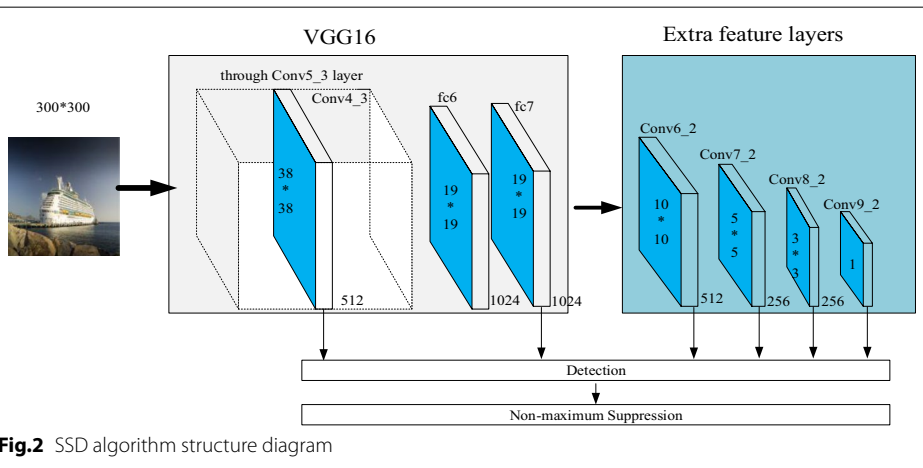

Fig.2 SSD algorithm structure diagram

(LFO-Net) based on popular single shot detector (SSD) model for single polarization SAR image ship detection.For ship detection, this method designed a simple lightweight network, proposing a bidirectional feature fusion module including semantic aggregation and feature reuse blocks, and used an attention mechanism to optimize features.

Compared with the above, the water surface target dataset constructed in this work contains five different targets and can be used for object classification. In addtion,this work uses a panoramic camera as a tool for object detection with a wider range compared to ordinary visual sensors. A comparison is shown in Table 1 below.

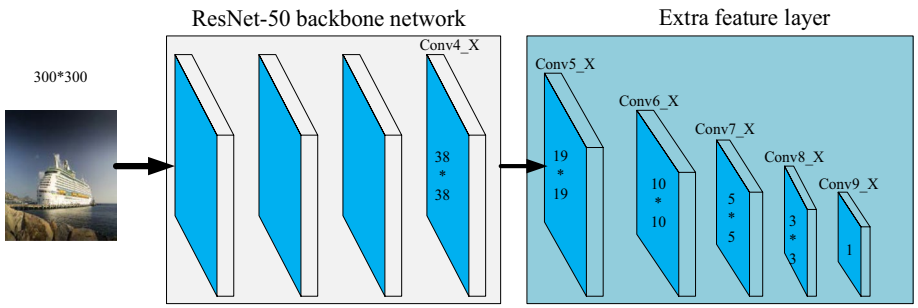

Fig. 3 The backbone structure diagram of the improved SSD network

Table 2 Backbone network parameters after modification

Name of convolutional layer	Input size	Kernel	Stride	Output size
Conv1_x	300*300	7*7, 64	2	150*150*64
Conv2_x	150*150*64	3*3 max pool	2	75*75*256
		$\begin{bmatrix} 1*1 & 64 \\ 3*3 & 64 \\ 1*1 & 256 \end{bmatrix}*3$	1	
Conv3_x	75*75*256	$\begin{bmatrix} 1*1 & 128 \\ 3*3 & 128 \\ 1*1 & 512 \end{bmatrix}*4$	2	38*38*512
Conv4_x	38*38*512	$\begin{bmatrix} 1*1 & 256 \\ 3*3 & 256 \\ 1*1 & 1024 \end{bmatrix}*6$	1	38*38*1024

3 Method

3.1 Panorama image preprocessing

The generation of a panoramic image relies on the processing of multiple photographs and mapping them on the geometric surface, forming an image by analyzing the picture seams and a seamless mosaic with image mosaic technology. Based on different projection modes, panoramic images can be classified as plane, cylindrical, spherical, and cubic. According to practical task requirements, this study adopts spherical projection.

A spherical panoramic image has a 360-degree horizontal view angle and 180-degree vertical view angle, and is shot by a fisheye lens. When identifying an object, three-dimensional spherical coordinates must be converted to two-dimensional plane coordinates. One common method is longitudinal and latitudinal mapping [29]. It is expanded based on the projection of spherical longitude and latitude on the surface of a circumscribed cylinder. The closer it gets to the two poles, the more apparent is the distortion. Set any point on the sphere as $M(\phi, \theta)$ and convert it to plane coordinates $N(x, y)$ through formula (1).

$$\begin{aligned} x &= \mathrm{W} * (\phi/2\pi) \\ y &= \mathrm{L} * (\theta + \pi/2)/\pi \end{aligned} \tag{1}$$

where ϕ and θ are the longitude and latitude, respectively, and W and L are the width and length, respectively, of the plane image. A rectangular plane graph with a length–width ratio of 2:1 is obtained, shown as Fig. 1.

3.2 Improve SSD object detection algorithm

The SSD algorithm is based on a feedforward neural network. The network structure shown as Fig. 2, takes a VGG16 network as the backbone, changing the last two fully connected layers to convolutional layers and adding four convolutional layers. There are six feature extraction layers with different scales. The SSD object detection algorithm has the following steps:

(1) Add convolutional layers with decreasing scales to the basic network, generate a multi-scale feature graph, and extract object features;
(2) Set prior boxes with different scales and length–width ratios for every pixel at the feature layer;
(3) Connect feature graphs of different sizes to the ultimate detection layer to position and classify objects;
(4) Calculate and output the result with the non-maximum suppression algorithm.

The SSD algorithm has advantages in terms of detection speed and precision, but suffers from poor detection of small objects, mainly because of insufficient extraction of deep features and limited information about small objects. The SSD algorithm detects objects of different sizes by utilizing feature graphs with different scales. Shallow and deep feature graphs are used to detect small and large objects, respectively. A shallow feature graph is characterized by a large mapping size and abundant details and features of objects, but it has poor semantic information. A deep feature graph has a small mapping size and abundant semantic and abstract information, but insufficient details and features of objects. It generally detects small objects through the conv4_3 layer at the lowest layer, but the precision is poor in this case. Therefore, the detection effect for small remote objects is unsatisfactory.

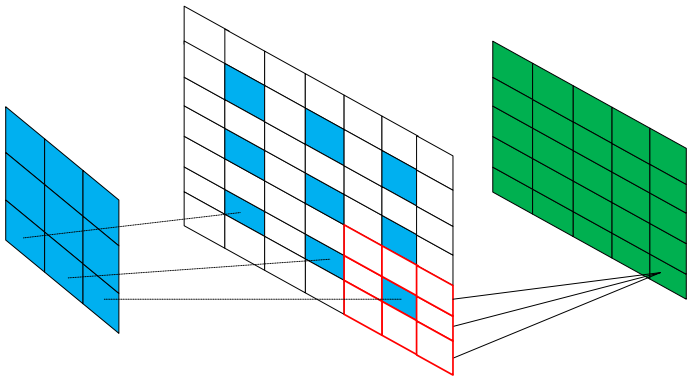

Fig. 4 Deconvolution operation

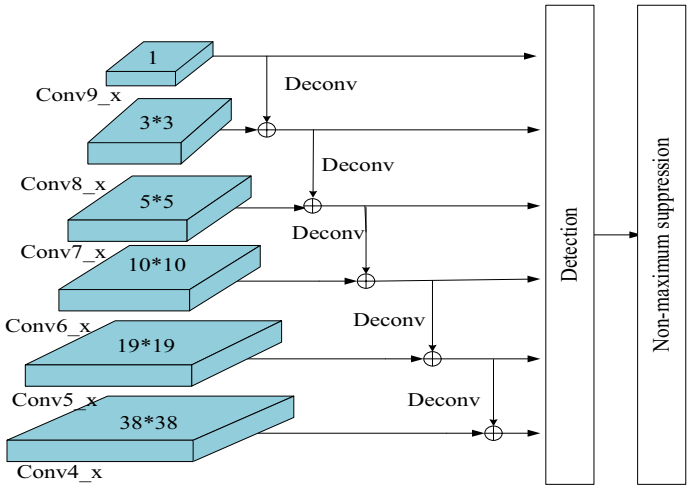

Fig.5 Feature pyramid network structure

Table 3 Deconvolutional operation parameters

Feature layer	Input size	Kernel	Stride	Padding	Output size
Conv9_x	1*1*256	3	2	0	3*3*256
Conv8_x	3*3*256	3	2	1	5*5*256
Conv7_x	5*5*256	2	2	0	10*10*256
Conv6_x	10*10*512	3	2	1	19*19*512
Conv5_x	19*19*512	2	2	0	38*38*1024

Fig. 6 Sample image of the dataset

3.2.1 Backbone network improvement

To obtain abundant semantic information of small objects, ResNet-50 [30] with residual learning units is used instead of VGG16 as the backbone network. ResNet-50 can solve the "vanishing gradient" problem caused by a deep network in the neural network. It has a deeper network than VGG16, can better extract feature graphs with more abundant semantic information, and has fewer parameters and a more prominent effect.

The improved backbone network is shown as Fig. 3. We set the conv4_x layer in the ResNet-50 network structure as the first feature extraction layer of SSD, remove the conv5_x and fully connected layers, and add five convolutional layers. Parameters for the backbone network after modification are shown in Table 2.

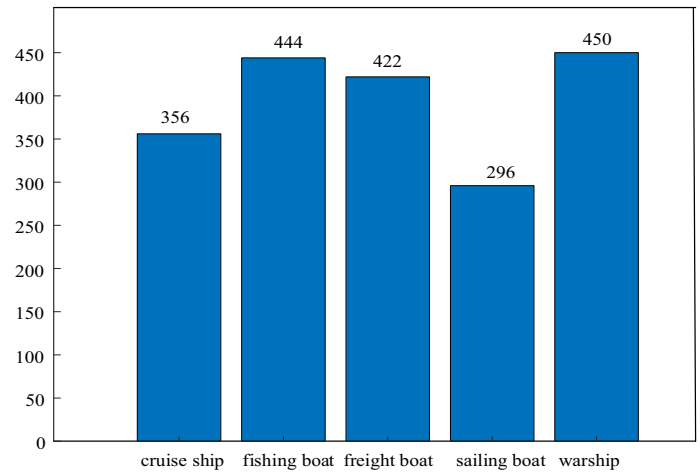

Fig. 7 Data distribution

3.2.2 Feature pyramid structure

A top-down feature pyramid network (FPN) [31] structure can be adopted to solve the problem of insufficient semantic information in the shallow feature graph of the SSD model and insufficient detailed information in the deep feature graph. It integrates information of feature graphs in different sizes to offer sufficient semantic information of the shallow feature graph and detailed information of the deep feature graph and improve the overall detection precision.

Deconvolution can be adopted during integration of feature graphs in different sizes. This is the reverse of convolution. Deconvolution is similar to upsampling methods such as bilinear, nearest neighbor, and area interpolation, and can obtain a feature graph of high dimension. We enlarge the size of the input image by zero fill between neighboring elements, and then carry out a convolution operation. The deconvolution equation is

$$D = S * (I - 1) + K - 2P \tag{2}$$

where S is the stride, K is the size of the convolution kernel, P is the padding size, I is the size of the input feature graph, and D is the size of the output feature graph. As shown in Fig. 4, a 5×5 feature graph is obtained from deconvolution of a 3×3 feature graph.

Hence the adoption of deconvolution for feature graphs at the intermediate and upper layers of the SSD feature pyramid can obtain feature graphs with higher dimensions. This is integrated with corresponding original feature graphs with consistent dimensions to integrate the shallow feature graph in the feature pyramid with deep semantic information. Figure 5 shows the network structure of the feature pyramid adopting the deconvolution operation. We carry out deconvolution for Conv9_x, Conv8_x, Conv7_x, Conv6_x, and Conv5_x, and integrate these with the next neighboring layer. The parameters are shown in Table 3. Six feature graphs of different sizes are generated after integration.

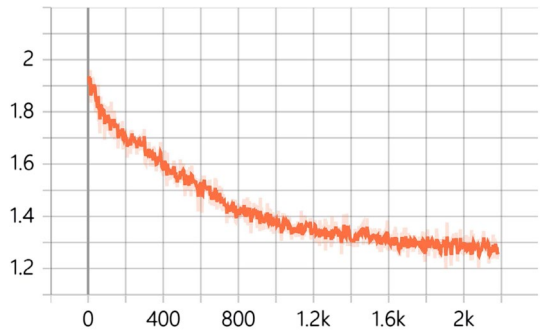

Fig. 8 Loss change

Table 4 Test Results

Classification	Precision(%)	Recall(%)	AP(%)
Cruise ship	83.10	80.73	86.96
Fishing boat	85.32	91.17	92.50
Freight boat	95.08	93.93	96.59
Sailing boat	90.50	75.13	82.01
Warship	95.34	91.90	96.55

3.3 Default boxes adjustment

The SSD algorithm locates objects of different sizes by setting a series of a default boxes of different scales on different layer feature maps. Suppose we want to use m feature maps for prediction. The scale of the default boxes for each feature map is computed as:

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m-1}(k-1), k \in (1, m) \tag{3}$$

where $S_{\min}$ is 0.2 and $S_{\max}$ is 0.9, meaning the lowest layer has a scale of 0.2 and the highest layer has a scale of 0.9. Through the analysis of data samples, in order to meet the requirements of small object size, we will adjust the values of $S_{\min}$ and $S_{\max}$ to 0.1 and 0.8.

4 Experimental analysis

4.1 Experimental data set and platform

To realize the quick detection of a water surface object, a dataset of network learning must be built. We built five common ship image datasets through connection, shooting, crawler search, and other methods, as shown in Fig. 6. We obtained 1757 images by expanding them through data enhancements such as flip and color adjustment. The distribution of target numbers for each type is shown in Fig. 7. All images were manually annotated with the LabelImg tool, and they adopted the VOC format. The hardware environment was an AMD Ryzen 7 4800H CPU and Nvidia GeForce RTX 2060 graphics card, and the software environment was a 64-bit Windows 10 operating system and TensorFlow deep learning framework.

Table 5 Results of different detection algorithms

method	Backbone	input	Cruise ship (%)	Fishing boat (%)	Freight boat (%)	Sailing boat (%)	Warship (%)	mAP (%)	FPS
Faster-RCNN(1)	ResNet-50	800* 800	**94.82**	90.30	91.58	80.49	88.18	89.07	8
SSD(2)	VGG-16	300* 300	86.03	87.45	91.50	80.83	86.43	86.45	54
SSD(3)	ResNet-50	300* 300	86.80	90.97	92.35	81.42	88.52	88.01	60
YOLOv3(4)	Darknet53	416* 416	90.73	88.69	90.98	**82.90**	91.50	88.96	45
Ours(5)	ResNet-50	300* 300	86.96	**92.50**	**96.59**	82.05	**96.55**	90.92	30

Bold indicates the largest value in the column and the asterisk stands for multiplication

4.2 Evaluation and training

Advantages and disadvantages of the performance of object detection models can be evaluated from the aspects of detection precision and speed. Detection speed has units of frames per second (FPS) as the evaluation index, i.e., the number of images the model can detect per second.

For single-category detection, we take precision, recall, and average precision as evaluation indexes for detection precision. These are defined as

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

where TP, FP, and FN are the numbers of accurately detected, wrongly detected, and undetected objects, respectively. The area of the graph encircled by the precision rate and recall rate is considered the average detection precision (AP) of this kind of object.

In the case of multiple categories of detection, detection precision generally takes the mean average precision,

$$mAP = \frac{1}{N} \sum AP \tag{6}$$

as an evaluation index, where N is the number of object categories in the dataset.

The entire model basically adopts training strategies of the original SSD, including data enhancement, dimension, and scale setting of the prior frame, loss function, and non-maximum suppression. The dataset is divided into training and testing sets in an 8:2 ratio, and 10% of the training set data is taken as a certification set. The training speed can be accelerated by using the trained ResNet50 pre-training weight and freezing the universal part of the backbone network based on transfer learning. We input the resolution ratio image of 300×300, and set the batch size as 16. The initial learning rate of training is 0.0005. We adopt the early stopping function to prevent the training from overfitting, and end the training when the loss value does not decrease after 500 times. The changing curve of the loss function during the training process is shown as Fig. 8.

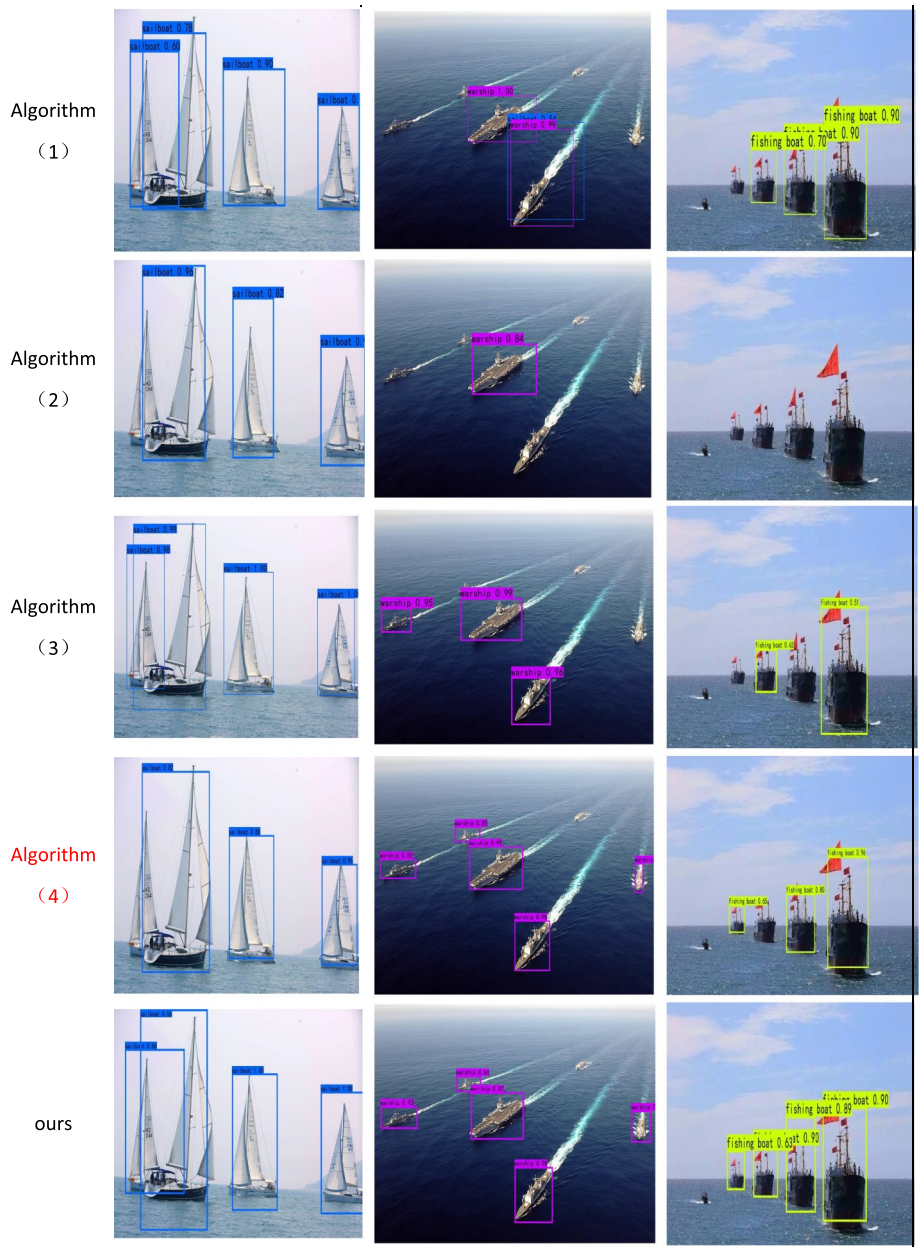

Fig. 9 Detection comparison between algorithms

There are 2,162 iterations in the entire training, each iteration took 62 s, and the loss value remains stable at 1.26.

5 Result analysis and discussion

To verify the performance of the proposed SSD algorithm, an experiment was conducted on the constructed water surface object dataset. The experimental results are shown as Table 4.

It can be seen from the results that the proposed algorithm showed good performance in detection of all kinds of objects. The average detection precision was over 80%, and it

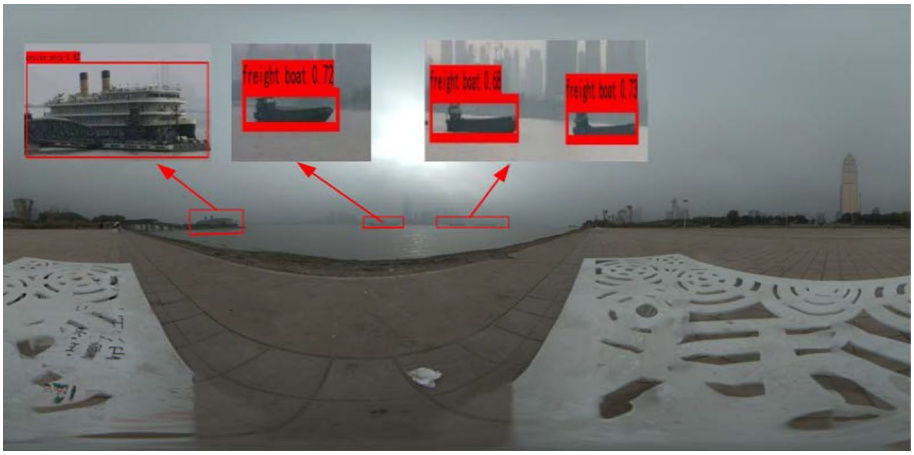

Fig.10 Detection result of panoramic visual picture

surpassed 90% for fishing boat, cargo ship, and warship, which demonstrates the effectiveness of this algorithm. The low accuracy of sailboats is mainly due to the small number of samples in the data set and the relatively small size of sailboats.

To further verify the performance of the algorithm, an experimental comparison was carried out with other object detection algorithms such as SSD, Faster-RCNN andY-OLOv3, with results as shown in Table 5.

It can be seen that the network structure is complicated, with the added FPN network and deconvolution operation, and the real-time performance of the algorithm was less than that of algorithms 2, 3 and 4, but the mAP increased by 3.8%, 2.3% and 1.6%, respectively. Compared with algorithm 1, the mAP increased by 1.2% and 1.6%, and the detection real-time performance of the proposed algorithm was clearly superior.

To show the advantages of the algorithm in this study, a picture with multiple objects for detection was selected. The detection result is shown in Fig. 9. It can be seen from Fig. 9 that algorithm 2 and 3 had missed detection for the picture on the left. All five algorithms had various degrees of missed detection for the middle picture, with the algorithm of this study and algorithm 4 having the lowest number of missed detection, and algorithm 1 also had erroneous detection. For the picture on the right, only the algorithm of this study could detect all ships. Through the experiments above, the improved SSD algorithm proposed in this study is seen to be superior to the other algorithms in terms of detection precision, especially for remote object detection.

To reflect the advantages of panoramic visual detection, we used the improved algorithm in this work for panoramic vision detection. The detection results are shown in Fig. 10. It can be seen from Fig. 10 that compared to ordinary visual pictures, panoramic visual pictures have a wider field of view. The target on the left side of the picture is a cruise ship, and the right side is three freight boats.

6 Conclusion

We carried out detection of five common ships on the sea surface based on the improved SSD algorithm and panoramic vision. More abundant object environmental information was obtained through the panoramic view. The reconstruction of the backbone network

with the advantages of Resnet50 improved the network depth, reduced the calculated number of parameters, and increased the object detection speed. We integrated feature graphs of different sizes and took full advantage of the semantic information of shallow feature graphs by integrating a feature pyramid network with deconvolution to improve the detection precision of remote objects. Experimental results revealed that the mean Average Precision (mAP) of the improved algorithm are increased by 4.03%, compared with the existing SSD detecting Algorithm, effectively reduce erroneous detection and missed detection of remote objects, and realize real-time detection. In addition, panoramic visual detection has a broader field of vision than ordinary visual detection, which can conduct global detection. We will next study methods to simplify the network structure so as to improve the detection speed, and expand the dataset to realize the detection of more objects.

Abbreviations
SSD: Single-shot Multibox Detector; CNN: Convolutional neural network; FPN: Feature pyramid network.

Acknowledgements
Not applicable

Authors' contributions
AL—editing, AL and XZ—experiments and analysis, AL and SH—data collection, JX—design conception. All authors read and approved the fnal manuscript.

Funding
The Natural Science Foundation of Hubei Province, China (Grant no.2018CFC865). The China Postdoctoral Science Foundation Funded Project (Grant no.2016T45686).

Availability of data and materials
The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Electronic Engineering, Naval University of Engineering, Wuhan 430000, China. [2]College of Armamente Engineering, Naval University of Engineering, Wuhan 430000, China.

Received: 7 September 2021 Accepted: 16 December 2021
Published online: 20 December 2021

References

1. D. Zhang, Y.X. Zhao, Y.F. Cui, P.C. Wang, A visualization analysis and development trend of intelligent ship studies. J. Transp. Inf. Saf. **39**(01), 7–16 (2021)
2. C.Q. Yi, Research on the development of panoramic image stitching technology. Inf. Comput. (Theoretical Edition), **32**(14), 149–151 (2020)
3. L. Meng, T. Hirayama, S. Oyanagi, Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2820326 (2018)
4. X. Ma, L.M. Shao, X. Jin, G.L. Xu, Advances in ship target recognition technology. Sci. Technol. Rev. **37**(24), 65–78 (2019). https://doi.org/10.3981/j.issn.1000-7857.2019.24.009
5. R. Wijnhoven, V.K. Rens, E. Jaspers, Online learning for ship detection in maritime surveillance, in *Proceedings of 31th Symposium on Information Theory in the Benelux Rotterdam the Netherlands 2010*, pp. 73–80.
6. S. Mirghasemi, H.S. Yazdi, M. Lotfizad, A target-based color space for sea target detection. Appl. Intell. **36**(4), 960–978 (2012)

7. T. Albrecht, G. West, T. Tan, et al. Visual maritime attention using multiple low-level features and Naive Bayes classification, in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications. IEEE Computer Society, 2011*, pp. 243–249.
8. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*. pp. 779–788.
9. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*, pp. 6517–6525
10. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement. arXiv arXiv:1804.02767 (2008).
11. A. Bochkovskiy, C.Y. Wang, O.H.Y.M. Lia, YOLOv4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
12. W. Liu, Anguelov D, Erhan D, et al. SSD: single shot multibox detector.Proceedings of the European Conference on Computer Vision, 2016: 21–37.https://doi.org/10.1007/978-3-319-46448-0_2.
13. R.Girshick, Fast R-CNN, in *2015 IEEE International Conference on Computer Vision; 2015 Dec 7–13; Santiago, Chile. Piscataway*: IEEEPress; 2015. pp. 1440–1448
14. S. Ren, K. He, R.B. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, CoRR abs/1506.01497. (2015) http://arxiv.org/abs/1506.01497.
15. J. Dai, Y. Li, K. He, et al. R-FCN: object detection via region-based fully convolutional networks. Neural Inf. Process. Syst. 379–387. (2016)
16. K. He, G. Gkioxari, P. Dollar, et al., Mask R-CNN, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017*. pp. 2980–2988.
17. X.F. Cao et al, Ship recognition method combined with image segmentation and deep learning feature extraction in video surveillance. Multimed. Tools Appl. https://doi.org/10.1007/s11042-018-7138-3 (2019)
18. L. Qi, B.Y. Li, L.K. Chen, Ship target detection algorithm based on improved Faster R-CNN. China Shipbuilding **61**(S1), 40–51 (2020)
19. Z. Lin, K. Ji, X. Leng, G. Kuang, Squeeze and excitation rank faster R-CNN for ship detection in SAR images. IEEE Geosci. Remote Sens. Lett. **16**, 751–755 (2019)
20. Q. Xiao, Y. Cheng, M. Xiao, J. Zhang, H. Shi, L. Niu, C. Ge, H. Lang, Improved region convolutional neural network for ship detection in multiresolution synthetic aperture radar images. Concurr. Comput. Pract. Exp. **32**, e5820 (2020)
21. D. Zhang, J. Zhan, L. Tan, Y. Gao, R. Župan, Comparison of two deep learning methods for ship target recognition with optical remotely sensed data. Neural Comput. Appl. **33**, 4639–4649 (2021)
22. Q.Z. Li, X.Y. Xu, Fast detection of surface ship targets based on improved YOLOV3-Tiny. Comput. Eng. (2021). https://doi.org/10.19678/j.issn.1000-3428.0059305
23. H.X. Xu, Z.S. Long, H. Feng, Multi-scale target detection algorithm for intelligent ship navigation. J. Huazhong Univ. Sci. Technol. (Natural Science Edition) (2021). https://doi.org/10.13245/j.hust.210509
24. Y. Li, J. Guo, X. Guo, K. Liu, W. Zhao, Y. Luo, Z. Wang, A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved YOLOV3. Sensors **20**, 4885 (2020)
25. Y. Jie, L. Leonidas, F. Mumtaz, M. Ali, Ship detection and tracking in Inland waterways using improved YOLOv3 and deepSORT. Symmetry **13**, 308 (2021)
26. Y.D. Li, H. Dong, H.G. Li, X.Y. Zhang, B.C. Zhang, Z.F. Xiao, Multi-block SSD based on small object detection for UAV railway scene surveillance. Chin. J. Aeronaut. **33**(6), 1747–1755. https://doi.org/10.1016/j.cja.2020.02.024 (2020)
27. Z.Y. Yin, C. Fan, Z.H. Zhao, Z. Huang, F.Q. Zhang, Target detection algorithm based on multi-scale feature map classification and re-extraction. Small Microcomput. Syst. **42**(03), 536–541 (2021)
28. X. Zhang, H. Wang, C. Xu, Y. Lv, C. Fu, H. Xiao, Y. He, A lightweight feature optimizing network for ship detection in SAR image. IEEE Access **7**, 141662–141678 (2019)
29. Y. Yang, X.R. Wang, Q. Dai, J.L. Fu, Research on spherical panoramic image generation technology. Comput. Appl. Softw. **24**(10), 164–187 (2007)
30. K. He, X. Zhang, S. Ren, S. Jian, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016*. pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
31. T. Lin, P. Dollár, R. Girshick, et al. Feature pyramid networks for object detection, in *Proceedings of the IEEE conference on computer vision and pattern recognition, 2017*. pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106

Publisher's Note